DOCUMENT RESUME

ED 451 189 SP 039 872

| | |
|---|---|
| AUTHOR | Brookhart, Susan M. |
| TITLE | The "Standards" and Classroom Assessment Research. |
| PUB DATE | 2001-03-02 |
| NOTE | 15p.; Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (53rd, Dallas, TX, March 1-4, 2001). |
| PUB TYPE | Information Analyses (070) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Academic Standards; Elementary Secondary Education; Grading; Standardized Tests; *Student Evaluation; Teacher Attitudes; *Teacher Competencies; Teachers |
| IDENTIFIERS | Teacher Knowledge |

ABSTRACT

This paper summarizes well-reviewed research about teacher competence in the skills described in the "Standards for Teacher Competence in the Educational Assessment of Students." It also examines additional research studies from 1990 to the present. The standards emphasize: choosing and developing assessment methods appropriate for instructional decisions; administering, scoring, and interpreting the results of externally produced and teacher produced assessment methods; using assessment results when making educational decisions; developing valid student grading procedures which use assessments; communicating assessment results to students, parents, and other lay audiences and educators; and recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of information. Three methods have been used to investigate teachers' knowledge and use of assessment: surveys of teacher attitudes, beliefs, and practices; tests of assessment knowledge; and reviews of teachers' assessments themselves. Teachers appear to do better at classroom applications than at interpreting standardized tests. They lack expertise in test construction, and they do not always use valid grading procedures. Few teachers provide criteria for grading or examples of student work ahead of time, and few involve students in creating criteria. Most studies have concluded that teachers need more instruction in assessment. (Contains 39 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 451 189

# The *Standards* and Classroom Assessment Research

Susan M. Brookhart
School of Education
Duquesne University
Pittsburgh, PA 15282
(412) 396-5190
brookhart@duq.edu

Paper presented at the annual meeting of the
American Association of Colleges for Teacher Education
Dallas, Texas
March 2, 2001

As part of the symposium
The Role of the *Standards for Teacher Competence in Educational Assessment Of Students* in Developing Quality Teacher Assessors

SP039872

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. M. Brookhart

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

## The *Standards* and Classroom Assessment Research

The *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA,1990) were a joint effort between the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. The introduction to the *Standards* states, "The standards should be incorporated into future teacher training [sic] and certification programs. Teachers who have not had the preparation these standards imply should have the opportunity and support to develop these competencies before the standards enter into the evaluation of these teachers." Thus the *Standards* acknowledge the importance of teacher education and professional development in assessment.

This paper summarizes research about teacher competence in the skills described in the *Standards*: "Teachers should be skilled in: (1) choosing assessment methods appropriate for instructional decisions; (2) developing assessment methods appropriate for instructional decisions; (3) administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; (4) using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement; (5) developing valid pupil grading procedures which use pupil assessments; (6) communicating assessment results to students, parents, other lay audiences, and [illegible]cators; and (7) recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information" (AFT, NCME, NEA, 1990).

All seven standards apply to teachers' development and use of classroom assessments of the various instructional goals and objectives that form the basis of classroom lessons. Standards 3, 4, 6, and 7 also include skills that apply to large scale assessment, including administering, interpreting, and communicating large-scale assessment results, using information for decision making, and recognizing unethical practices.

A body of literature addressing teacher preparation and teacher competence in assessment does exist (Table 1). In 1989, the Buros Institute held a symposium titled, "Are our school teachers adequately trained in measurement and assessment skills?" This resulted in an edited volume of papers (Wise, 1993) on the topic of what preservice teachers are taught about measurement, what they should be taught, what they actually know, and the observed quality of their assessment practices. Two chapters include literature reviews: Gullickson's (1993) review of literature on teachers' attitudes and practices and the content of undergraduate measurement courses, and Marso and Pigge's (1993) review of literature on teachers attitudes and practices and studies of direct assessments of teachers' testing knowledge.

Additional often-cited basic literature on this topic include an article by Schafer and Lissitz (1987) about what school personnel should know, do know, and have the opportunity to learn; an entire special issue of *Educational Measurement: Issues and Practice* (Nitko, 1991) on the topic; Gullickson's (2000) review of the need for Student Evaluation Standards; and Stiggins and Conklin's (1992) book about teacher practices in classroom assessment. One study was specifically based on the *Standards*: Plake, Impara, and Fager (1993) did a national survey, supported by a Kellogg Foundation grant to the National Council on Measurement in Education,

specifically to measure teachers' competency levels in the areas addressed by the *Standards* and then to develop an educational prototype based on an identified area of need.

**Table 1.**
**Selected Basic References about Teacher Preparation and Teacher Competence in Assessment**

| Reference | Title |
|---|---|
| Gullickson (1993) | Matching measurement instruction to classroom-based evaluation: perceived discrepancies, needs, and challenges |
| Gullickson (2000) | The need for student evaluation standards |
| Marso & Pigge (1993) | Teachers' testing knowledge, skills, and practices |
| Nitko (Ed.,1991) | What are we teaching teachers about assessment and why? |
| Plake, Impara, & Fager (1993) | Assessment competencies of teachers: A national survey |
| Schafer & Lissitz (1987) | Measurement training for school personnel: Recommendations and reality |
| Stiggins & Conklin (1992) | In teachers' hands: Investigating the practices of classroom assessment |

This list is not meant to be exhaustive; rather, it is intended as a recommended "starting point" for reading in this area. Since prior literature has been well reviewed, this paper will discuss the findings of those reviews and add individual research studies from 1990 to the present. Additional studies were found in two ways: (1) an ERIC search using the terms "teacher and (classroom assessment)" ; (2) e-mail request to members of the American Educational Research Association's Special Interest Group in Classroom Assessment in order to identify relevant conference papers and other fugitive sources.

## Teacher Preparation and Teacher Competence in Assessment

Many authors argue for an increase in emphasis on classroom assessment in teacher preparation and for a corresponding decrease in emphasis on large-scale testing (Gullickson, 1993; Marso & Pigge, 1993; Schafer, 1991, 1993; Stiggins, 1991a,b). Studies have generally concluded that teachers' knowledge and skills regarding both classroom assessment and large-scale testing are limited. While one might question the "what's wrong with this picture" approach to studying teachers' assessment competence and needs, that is the approach much of the literature has taken. Some of the more recent studies have also suggested ways to document improvement (e.g., Ashbacher, 1999). It is important to remember that while the focus of the *Standards* is "teacher competence" at assessment tasks, practicing teachers create a "classroom assessment environment" (Stiggins & Conklin, 1992) or "evaluative culture" (Haydel, Oescher, Kirby, & Brooks, 1997) that encompasses their assessment attitudes and values as well as their assessment practices, and their integration of those with their instructional and classroom management practices.

Knowledge and Use of Assessments

Three methods have been used to investigate teachers' knowledge and use of assessment: surveys of teacher attitudes, beliefs, and practices; tests of assessment knowledge; and reviews of teachers' assessments themselves. Table 2 gives a very generalized summary of the results of this body of research and a sample of references. Most studies have concluded that teachers need more instruction in assessment (Marso & Pigge, 1993).

**Table 2.**
**The *Standards* and Classroom Assessment Research**

| Standard | Summary of Findings |
|---|---|
| 1. Choosing assessment methods appropriate for instructional decisions | Some teachers do; many don't (Aschbacher, 1999; Marso & Pigge, 1993) Informal assessment is much-used but little-discussed and requires more attention (Oosterhof, 1995) |
| 2. Developing assessment methods appropriate for instructional decisions | Some teachers do; many don't (Aschbacher, 1999; Haydel, Oescher, & Banbury, 1995; Marso & Pigge, 1993) |
| 3. Administering, scoring, & interpreting the results of both externally-produced and teacher-produced assessment methods | Teachers are better at administration of tests than communicating results (Plake, Impara, & Fager, 1993) Scoring experience can make teachers better classroom assessors (Goldberg & Roswell, 1998) |
| 4. Using assessment results when making educational decisions | Using assessment results is difficult for teachers (Zhang, 1996) Tension between teacher beliefs & values and classroom realities & external factors influences teacher decision making (McMillan & Nash, 2000) |
| 5. Developing valid pupil grading procedures which use assessments | Some teachers do; many don't; The treatment of student effort is a particular issue (Brookhart, 1994; McMillan, in press) Teachers' beliefs and philosophy of education are important (McMillan & Nash, 2000) |
| 6. Communicating assessment results to students, parents, other lay audiences, & other educators | Many teachers have problems with this (Imapra, Divine, Bruce, Liverman, & Gay,1991; Plake, Impara, & Fager, 1993) |
| 7. Skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of information | Teachers are better at recognizing unethical practices than at scoring or interpreting results (Plake, Impara, & Fager, 1993) |

Gullickson (1993) found that a large proportion of teachers reported using classroom tests: 89% of elementary and 99% of secondary teachers. More frequent, shorter tests were preferred to longer tests. Most of the classroom tests teachers used, they reported, were teacher-made tests

(86% of elementary, 97% of junior high, and 96% of senior high teachers reported using teacher made tests) and tests from textbook publishers (75% elementary, 61% junior high, 47% senior high). Multiple choice and essay questions were the most commonly reported item formats. Gullickson (1993) also reported that most teachers did not support using test scores as the sole determinant of students' report card grades.

Marso and Pigge (1993) reviewed studies, including their own, investigating teacher's knowledge of tests as indicated by testing the teachers themselves or reviewing actual classroom tests. They concluded that while teachers did not score well on tests about classroom testing, those reporting measurement instruction only scored 6% to 10% better than those not reporting measurement instruction, illustrating either poor quality instruction or a disconnect between the conventional instruction of a decade ago and teachers' classroom needs. Reviews of classroom tests found frequent use of objective test items and 70% to 100% of the items at the recall level, with math and science tests faring a little better than other subjects. Teachers did not follow the principles of good item writing and were especially poor at constructing matching exercises and at putting clear directions on tests. On a more positive note, teachers' and principals' ratings of their own proficiency at classroom test construction were negatively related to the amount of errors: the more proficient educators thought they were, the fewer errors they made, thus documenting at least some measure of self-awareness regarding classroom test construction.

However, self-awareness about test construction practices may be partial. Oescher and Kirby (1990) examined classroom tests of 19 science and math teachers, finding as expected many recall items, but noting that the classroom teachers' judgments regarding the cognitive demand of their items was often above that of researchers' judgments. Teachers categorized about a quarter of their test items as testing higher-order thinking, while researchers categorized about 8% as such. Oescher and Kirby (1990) pointed out that teachers may discount the influence of instruction, which can change what seems to be a higher-order question into recall – of a discussion, or a section of a textbook, for instance.

Oosterhof (1995) made extended observations of 15 classroom teachers. He reported that teachers spent more time on informal assessment for ongoing monitoring of classroom activities than on formal assessment such as giving a test. Nevertheless, he pointed out, most measurement texts used in teacher education do not put much emphasis on informal assessment. Nor do the *Standards*. Standard one, choosing appropriate assessments, mentions informal assessments, and some of the other standards use language that could be interpreted to include informal assessments, but in the main the *Standards* discuss formal assessment procedures.

Aschbacher (1999) and her colleagues at CRESST have begun a project to develop indicators of classroom practices based on samples of assignments and student work. The purpose of this work is to develop indicators capable of monitoring the results of assessment reform efforts as they play out in the classroom. The main purpose of Aschbacher's study was to investigate the indicator system itself, but its results are instructive in that they accord with the previous research about classroom assessment quality. The indicators, however, are much more closely tied to real classroom contexts and practices than the previous research and thus represent an important development in classroom assessment research. Based on assignments, interviews, and student work, Aschbacher (1999) developed a set of evaluative rubrics: cognitive demands

of the task, clarity of grading, alignment of task with learning goals, alignment of criteria for grading with learning goals. Designating teachers whose assessments demonstrated alignment of both task and grading criteria with learning goals as having "coherent" assessments, the study found that only one quarter to one third of middle school teachers in the sample and one quarter to half of the elementary teachers in the sample had coherent assessments (Aschbacher, 1999).

Haydel, Oescher, and Banbury (1995) examined a sample of performance assessments from elementary and middle school teachers in one district as part of a district effort to monitor its emphasis on complex outcomes for students. Haydel and her colleagues identified the most problematic areas in these teacher-developed performance assessments: poorly articulated purpose, lack of comprehensiveness and specificity (e.g., "be creative"), and lack of description of quality on the rubrics or quality continuum used.

Cox (1995) surveyed 155 high school English teachers in California. These high school English teachers reported the highest levels of routine use for daily homework or class assignments and teacher-developed tests or quizzes. Group or class participation was the only non-writing, performance activity indicated as routine by a majority (64%) of teachers. Writing assignments used routinely for assessment included final drafts of writing assignments, stages of the writing process, and integrated reading/writing assessments (a type of assessment used in the California state assessment system in 1993). Non-achievement factors routinely considered in assessment included completion of assignments and on-time submission of assignments. Cox (1995) also addressed questions about the assessment reform goal of educational opportunity to the same survey data. She found that honors classes received more writing assignments than classes with average or below average students, and classes with few minority students used portfolios more often than classes with many minority students.

Snow-Renner (1998) reported the results of a study of Colorado mathematics teachers designed to investigate classroom assessment practices relative to Colorado's content, performance, and opportunity-to-learn standards. Both elementary and secondary teachers reported spending 9% to 11% of their instructional time on testing, both classroom and standardized. Additional questions asked about nature of these assignments and tests. Both elementary and secondary teachers reported that 43% of their tests were performance-based. Forty percent of elementary and 33% of secondary teachers reported that their assignments required students to apply concepts to new situations or problems. There was, however, much variability in the data; standard deviations for these averages were very large. Snow-Renner concluded that opportunity to learn and opportunity to show reform-relevant mathematics skill varied widely from classroom to classroom.

Mertler (2000) surveyed 625 teachers in Ohio specifically about the methods they used to guarantee the reliability and validity of the classroom assessments they used. He found that teachers had a better grasp of reliability than validity, but that practices to insure quality classroom assessment information in general were inadequate.

Plake, Impara, and Fager (1993) reported the results of a national study of teacher competencies related to the *Standards* and found that regarding large-scale assessment, teachers were more successful at questions about administering tests and about recognizing unethical practices than

they were at interpreting and communicating results. This accords with prior studies of teachers' knowledge of large-scale assessment concepts (Marso and Pigge, 1993). In a study in the state of Virginia, Impara and his colleagues (Impara, Divine, Bruce, Liverman, & Gay,1991) found that teachers also had problems interpreting test score reports, although providing interpretive information on printouts helped.

The Plake, Impara, and Fager (1993) study is particularly important to the topic of this paper for two reasons. First, the survey itself consisted of two parts: a test based specifically on the *Standards*, and a survey of teachers' beliefs and perceptions about test use. Each Standard was the basis for a 5-item subscale on the test. Table 3 presents the means (and standard deviations) for 555 teachers in this national sample in descending order. Second, the study used a national sample and thus yielded more generalizable results than many of the other studies reviewed here.

**Table 3.**
**Results for a National Sample of 555 Teachers (Plake, Impara, & Fager, 1993, p. 12) on 5-Item Subtests Based on the *Standards***

| Standard | Mean (Standard Deviation) |
|---|---|
| 3. Administering, scoring, & interpreting | 3.96 (.90) |
| 1. Choosing assessment methods for instructional decisions | 3.46 (.93) |
| 4. Using assessment results for decisions | 3.40 (1.11) |
| 7. Recognizing unethical, illegal methods and uses | 3.26 (.78) |
| 2. Developing assessment methods for instructional decisions | 3.22 (.80) |
| 5. Developing valid pupil grading procedures | 3.19 (.78) |
| 6. Communicating assessment results | 2.70 (1.21) |

Of the 555 teachers in their sample, nearly 70% had had some training in measurement, either preservice or inservice. While this instruction made a statistically significant difference in performance on the test based on the *Standards*, that difference was less than one point on the 35-point test: The average score for those who had taken a measurement class was 23.41, while the average score for those who had not was 22.72 (Plake, Impara, & Fager, 1993, p. 12). This accords with the finding in Marso and Pigge's (1993) review that measurement instruction does not make as much of a difference as one would hope.

Goldberg and Roswell (1998) surveyed teachers who had participated in scoring the Maryland State Performance Assessment (MSPAP), and followed up with interviews of selected teachers. Teachers expressed that their participation in scoring helped them become more thoughtful about their own classroom assessment in general, and performance assessment in particular. They reported that after their scoring experience, their own classroom assessments were more likely to be coherent with classroom learning goals and to elicit higher order thinking.

Zhang (1996) found that a sample of teachers' self-ratings on an Assessment Practices Inventory developed in part from the *Standards* yielded six factors. She ranked these factors according to teachers' perceptions, from most to least difficult. Interpreting standardized test results and using assessment results in decision making was the most difficult factor; communicating assessment results was the least difficult. This is directly contrary to the national data reported by Plake, Impara, and Fager (1993) and may be related to differences in sampling. Developing

and using performance assessments and informal observations was perceived as more difficult than developing and using paper and pencil tests.

Brookhart (1994) reviewed 19 studies of teachers' grading practices. She found that teachers try hard to be fair to students, and one important part of being "fair" is perceived as informing students ahead of time what will count in a grade. This does contribute to valid grading practices. A second finding matches less well with the standard: achievement measures are the major components teachers consider when assigning grades, but effort and ability are also commonly considered. Grading practices vary by grade level, with elementary teachers relying more on informal evidence, and also vary considerably by individual teacher (Brookhart, 1994).

McMillan (in press) surveyed 1483 secondary (6-12) classroom teachers of academic subjects in Virginia about their grading practices. Academic achievement was the most often cited component of grades, but "academic enablers" (effort, ability, improvement, participation) were also sometimes important components of grades. As grade level increased, reliance on publishers' tests diminished and use of teacher-made tests increased. This study replicated the results of earlier grading practices surveys, but with recent data in a large-scale survey, establishing that these conclusions are still current and have not changed with the advent of the educational accountability movement and state testing programs.

McMillan and Nash (2000) interviewed 24 teachers, selected because they represented a wide range of responses on the grading practices survey, about the reasons for their assessment and grading decisions and their use of assessment information in educational decisions, including decisions about what grades to assign. The most important internal factors affecting grading decisions appeared to be the teachers' philosophy of teaching and learning, the value they place on pulling for students, and their beliefs about promoting student progress and accommodating individual differences. Classroom realities and external factors such as student home environment, district policies and programs, absenteeism, and the like affected the rationales given for assessment and grading decisions, too. Often these external factors were in tension with the teachers' intentions and dispositions about promoting student learning. Teachers' focus was on promoting the learning of individual students, not necessarily creating assessments or grades that were comparable or consistent measures across students.

Instruction about assessment

Several approaches have been taken to the question of what teachers and prospective teachers should be taught about assessment. These have included surveys of teachers and of measurement experts, reviews of program and course requirements, reviews of textbook contents, and observation and job analysis of teaching. This literature has been well reviewed (see especially Gullickson, 1993). Generally, surveys of teachers and the job-analysis approach have concluded by recommending more instruction, at both the preservice and inservice level, to build a repertoire of strategies for high-quality classroom assessment and less instruction in standardized testing than is currently the case in most measurement courses for teachers. Textbook reviews or surveys of measurement experts have placed more emphasis on learning about standardized testing, although some measurement experts who have studied the classroom and its information needs do not (Airasian, 1991). Recommendations often include a shift away

from statistical interpretation of test scores and toward communication of score meaning with a lay audience (Brookhart, 1999; Schafer, 1991; Schafer & Lissitz, 1987). The relatively low level of communication competence reported in the literature adds empirical support to this recommendation.

Gullickson (1993) summarized the results of several survey studies he and his colleagues conducted in the early 1980's. He compared the 20 content priorities rated most highly by teachers and professors, respectively. Both groups emphasized the importance of teaching about the preparation of classroom tests. Professors reported more coverage of test statistics and analysis and of standardized test applications than teachers desired, and teachers desired more topics covered under formative and summative uses of tests and nontest evaluation practices than professors reported (Gullickson, 1993, p. 12-13).

Several authors present examples of high-quality measurement instruction and experiences for teachers and empirical evidence that they work. Taylor and Nolen (1996) demonstrated that an assessment course for preservice teachers can be based on a thoughtful connection between assessment principles and the classroom context, and that the result was a deeper understanding of assessment and a deeper appreciation of its value for teaching.

Burry-Stock, Schaffner, and Cho (1999) reported on an undergraduate educational measurement course aligned with the *Standards* and with considerations of classroom work expectations. Their students perceived their assessment competencies to be greater than colleagues in their state who attended other institutions. Their report (Burry-Stock, Schaffner, & Cho,1999) includes a syllabus and description of the course.

O'Sullivan and Johnson (1993) described a graduate-level educational measurement course that incorporated performance tasks built around the *Standards*. Their results indicated students did increase their assessment competencies and agreed that course objectives based on the *Standards* had been met. O'Sullivan and Johnson's (1993) paper includes a brief description of the course and eight performance tasks.

Plake, Impara, and Wise (1997) designed materials for teachers to be resources for teacher professional development. Their national survey of teacher competence on the *Standards* (Plake, Impara, & Fager, 1993) had identified interpreting and communicating assessment results as an area of need; therefore, this was the topic area for the materials. Materials were field tested with both preservice and inservice teachers and met with some success. Suggestions for improvement were reported, as well.

## Discussion

Any discussion of this body of research must begin by acknowledging a couple of major limitations to the research. The major limitation is sampling; many of the studies have been convenience samples in one geographic area, and all have been based on voluntary participation. A second limitation has been the design of many of the research questions from the point of view of measurement principles only, without consideration of the principles of instruction and

classroom management (Brookhart, 1999); this limits the fidelity of the questions to the practice of teaching.

Nevertheless, some consistencies occur, no matter what the sample, in the research on classroom assessment. These have been summarized in Table 2 and described in this paper. Teachers apparently do better at classroom applications than at interpreting standardized tests, as might be expected from the nature of their work. They lack expertise in test construction. They do not always use valid grading practices. This author wonders, however, whether testing teachers about educational psychology, or human development, or any of the other theoretical bases for education, result in any greater pronouncements of "competence"?

Gullickson (1993) pointed out that measurement experts have written about what teachers need to know about assessment since at least the first volume of the *Journal of Educational Measurement* (Mayo, 1964). Four decades of studies of "teacher competence" have not succeeded in raising "competence" to "standard," at least in the sense of large numbers of teachers reporting comfort with and skill at the assessment process. But "competence" at these standards, in the narrow sense, is not enough. And Marso and Pigge (1993) have pointed out that this "teacher competence" approach has resulted in measurement professionals advocating "a relatively undirected encouragement of better training practices, of further research of the problem, and of communications describing the problem" (p. 174).

Stiggins (1999) suggested a different set of competencies, some of which match the competencies in the *Standards* but are organized differently or are more detailed. Two competencies on Stiggins's list that are not in the *Standards* are: (a) "Start the classroom assessment development process with a highly refined vision of the specific academic achievements they want their students to attain;" and (b) "Define assessment as a student-centered rather than just a teacher-centered activity by involving students deeply in the assessment, record keeping, and communication process" (Stiggins, 1999, p. 24).

Equipping teachers with a highly refined vision of the specific academic achievements they intend for their students is part of the motivation behind the movement in teacher education to require more academic coursework for teachers. Many programs are exploring the addition of cognate areas of study for education majors, for example. Some schools are redesigning their teacher education programs as masters level programs, which student will enter after receiving a bachelor's degree in an academic major. The professional development work of Stiggins and others, raising awareness of the central importance of the achievement target, is helpful for practicing teachers. Goldberg and Roswell's (1998) results also indicated that focused training and practice in scoring performance assessments had the effect of highlighting for teachers the importance of the match between assessments and achievement targets.

Ferrara, Goldberg, and McTighe (1995) investigated how teachers communicate their expectations to students – how they describe their learning targets and standards of performance quality to students. Teachers most often communicated their expectations orally (36%) or by demonstration (38%). The most common way to do this was to state outcomes or objectives (69%) at the start of a lesson (58%). Few teachers provided criteria for grading or examples of student work ahead of time, and few invovled students in the creation of those criteria.

BEST COPY AVAILABLE

Student-centered assessment could be a model for teacher preparation as well as for the classroom. It might be productive to involve teachers in this quest for improvement of practice (Gallagher, 2000) in the same way it is recommended that all students be involved in their own assessment (Stiggins, 1999). Many teacher education programs are already involving teachers in their design. In order for teachers to own the goal of high-quality assessment, they must be convinced that assessment is just as important to student learning as is, say, effective classroom management or lesson design. Teacher preparation programs can help that happen by offering opportunities for preservice and practicing teachers to explore classroom asssessment, emphasizing the direct connection of clear achievement targets and student involvement in assessment (Stiggins, 1999) with student learning.

## References


Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational measurement: Issues and Practice, 10*(1), 13-16, 26.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: Authors.

Aschbacher, P. R. (1999, December). Developing indicators of classroom practice to monitor and support school reform. University of California, Los Angeles: CRESST Technical Report 513.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*, 279-301.

Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational measurement: Issues and Practice, 18*(1), 5-13.

Burry-Stock, J. A., Schaffner, M., & Cho, G-P. (1999, April). *Using a standards-driven curriculum in teaching classroom assessment: Making a difference in teachers' perceived skill competency*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Cox, K. B. (1995). *What counts in English class? Selected findings from a statewide study of California high school teachers*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Number ED 384 051)

Ferrara, S., Goldberg, G., & McTighe, J. (1995). *Ways in which teachers communicate learning targets, criteria,and standards for performance to their students*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Number ED 386 436)

Gallagher, C. (2000). A seat at the table: Teachers reclaiming assessment through rethinking accountability. *Phi Delta Kappan, 81*, 502-507.

Goldberg, G. L., & Roswell, B. S. (1998). *Perception and practice: The impact of teachers' scoring experience on performance-based instruction and classroom assessment*. Paper presented at the annual meeting of the American Educational Research Association, San Diego. (ERIC Document Number ED 420 670)

Gullickson, A. R. (1993). Matching measurement instruction to classroom-based evaluation: perceived discrepancies, needs, and challenges. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1-25). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Gullickson, A. R. (2000, May). *The need for student evaluation standards*. Western Michigan University: Joint Committee on Standards for Educational Evaluation. Available http://www.wmich.edu/evalctr/jc/SESNeed.html

Haydel, J. B., Oescher, J., & Banbury, M. (1995, April). *Assessing classroom teachers' performance assessments*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Haydel, J. B., Oescher, J., & Kirby, P. C. (1999). *Relationships between evaluative culture of classrooms, teacher efficacy, and student efficacy*. Paper presented at the annual meeting of the American Educational Research Association,Montreal.


BEST COPY AVAILABLE

Haydel, J. B., Oescher, J., Kirby, P. C., & Brooks, C. R. (1997, March). *Measuring the evaluative culture of classrooms*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10*(4), 16-18.

Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Teachers' assessment background and attitude toward testing. *Theory into Practice, 32*, 113-117.

Marso, R. N., & Pigge, F. L. (1993). Teachers' testing knowledge, skills, and practices. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 129-185). Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Mayo, S. T. (1964). What experts think teachers ought to know about educational measurement. *Journal of Educational Measurement*, 1, 79-86.

McMillan, J. H. (in press). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice.*

McMillan, J. H., & Nash, S. (2000, April). *Teacher classroom assessment and grading practice decision making*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Mertler, C. A. (2000). Teacher-centered fallacies of classroom assessment reliability and validity. *Mid-Western Educational Researcher, 13*(4), 29-35.

Nitko, A. J. (1991). Entire issue -- What are we teaching teachers about assessment and why? *Educational Measurement: Issues and Practice, 10*(1), 2.

Oescher, J., & Kirby, P. C. (1990). *Assessing teacher-made tests in secondary math and science classrooms*. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston. (ERIC Document Number ED 322 169)

Oosterhof, A. (1995). *An extended observation of assessment procedures used by selected public school teachers*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Number ED 390 937)

O'Sullivan, R. G., & Johnson, R. L. (1993). *Using performance assessment to measure teachers' competence in classroom assessment*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta. (ERIC Document Number ED 358 156)

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10-12, 39.

Plake, B. S., Impara, J. C., & Wise, V. L. (1997). Development and validation of professional Development resource materials for teachers covering communicating and interpreting assessment results. *Educational Measurement: Issues and Practice, 16*(2), 19-24.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice, 10*(1), 2-6, 12.

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice, 32*, 118-126.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57-63.

Snow-Renner, R. (1998). *Mathematics assessment practices in Colorado classrooms: Implications about variations in capacity and students' opportunities to learn.* Paper presented at the annual meeting of the American Educational Research Association, San Diego. (ERIC Document Number ED 424 293)

Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan, 72*, 534-539.

Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7-12.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.

Taylor, C. S., & Nolen, S. B. (1996). What does the psychometrician's classroom look like?: Reframing assessment concepts in the context of learning. *Education Policy Analysis Archives, 4*(17). Available http://olam.ed.asu.edu/epaa/v4n17.html

Wise, S. L. (Ed.). (1993). *Teacher training in measurement and assessment skills.* Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska-Lincoln.

Zhang, Z. (1996). *Teacher assessment competency: A Rasch model analysis.* Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Number ED 400 322)